ChinaXiv合作期刊
图书情报工作
第64卷 第11期 2020年6月

新兴技术主题识别方法研究进展*

■ 刘小玲[1,2] 谭宗颖[1]

[1] 中国科学院文献情报中心 北京 100190 [2] 中国科学院大学经济与管理学院图书情报与档案管理系 北京 100190

摘 要：[目的/意义]新兴技术主题识别不仅有助于及时跟踪技术发展动态，更能尽早捕捉技术领域未来的发展契机和可能的变化趋势。梳理新兴技术主题识别的定量研究方法，并对其优缺点进行比较，可以为新兴技术主题识别方法的改进和完善提供参考。[方法/过程]首先对"新兴技术""新兴技术主题识别"等概念的内涵进行辨析；然后调研和系统梳理国内外新兴技术主题识别的定性和定量研究方法，重点关注以文献计量和数据挖掘为主的定量研究方法，并将其划分为三类：主题词或文献统计方法、引文网络聚类方法和文本挖掘分析方法；最后综合分析各类研究方法在技术主题抽取、新兴技术主题识别指标体系构建、方法有效性验证等方面的异同和存在的缺陷，以及对方法改进的初步思考。[结果/结论]三类方法在新兴技术主题识别的主要步骤上各有特点和优劣，均有进一步完善的空间，未来可以探索利用深度学习等技术进行技术主题的准确抽取，并构建更加全面、系统的新兴技术主题识别指标体系，以及基于机器学习进行更加严格的方法有效性验证。

关键词：新兴技术 主题识别 方法研究 文献计量 文本挖掘

分类号：G306.0

DOI：10.13266/j.issn.0252-3116.2020.11.016

1 引言

随着社会发展和科学技术进步，建立在信息技术、生物技术和其他学科基础之上的新兴技术不断出现和快速发展，这些技术的出现与发展不仅改变了传统产业的发展形态，而且改变了人们的意识、观念、生活方式以及社会经济生产方式，对人类社会的影响越来越显著和深远。监测全球技术前沿变化、识别新兴技术不仅能及时跟踪技术发展动态，更能尽早捕捉未来的发展契机和可能的变化趋势，这对一个国家或地区的未来发展至关重要。同时，新兴技术识别还可以为政策制定者、企业、研究机构和研究人员提供科技研究趋势和优先技术领域的变化方面的信息，从而为决策者寻找资助的技术领域和资助对象提供支撑；帮助企业和研究机构决定其未来的研究定位和优先领域，识别潜在的合作对象；帮助研究人员及时了解技术领域发展的新动向；帮助投资机构通过对新兴技术领域和关键创新者的早期投资而获得回报。

自20世纪90年代中期"新兴技术"概念提出以来，学者们就如何寻找、发现和识别新兴技术开展了大量的研究工作，基于文献计量和数据挖掘的定量分析是常用的研究方法。本文在梳理这些研究方法的基础上对其优缺点进行比较，以期为改进和完善新兴技术主题识别方法提供参考。

2 相关概念

2.1 "新兴技术"及相关概念

20世纪90年代中期，美国宾夕法尼亚大学沃顿商学院的"新兴技术管理研究计划"首次提出"新兴技术(Emerging Technology)"概念。沃顿商学院的研究人员认为，新兴技术指"建立在科学基础上的，可能创立一个新行业或改变某个老行业的创新"，他们认为新兴技术不仅包括产生于根本性创新的技术，例如生物制药、数字成像、高温超导体、微型机器人和笔记本电脑等，还包括通过集成多个过去独立的研究成果而更具创新意义的技术，例如核磁共振成像、传真、电子金融和互联网等技术[1]。"新兴技术"概念提出后，国内研究人员鲁若愚[2]、李仕明[3]、徐建国[4]等以及国外研究人员S. Cozzens[5]、A. Breitzman等[6]给出了不同定义(见表

* 本文系国家自然科学基金项目"科学基金学科演变及资助政策研究"(项目编号:71843005)研究成果之一。

作者简介：刘小玲(ORCID:0000-0001-7523-247X)，助理研究员，博士研究生；谭宗颖(ORCID:0000-0003-3945-7174)，研究员，博士生导师，通讯作者，E-mail:tanzy@mail.las.ac.cn。

收稿日期:2019-04-18 修回日期:2019-11-17 本文起止页码:145-152 本文责任编辑:杜杏叶

1)，如鲁若愚等[2]认为“新兴技术是一种新观念、新方法、新发明，它以科学为基础，并能够创造一个新的行业或者能够改变一个现有行业并能对经济结构产生重大影响”，S. Cozzens 等[5]在分析了近 2 000 篇涉及新兴技术文章的基础上，将新兴技术定义为“是快速增长、新兴、具有未开发的市场潜力和高科技基础的技术，这种技术有巨大的潜力，但尚未表现出价值或在业界达成共识”，他认为新兴技术具有以下四个特征：快速增长；在转变过程中或变为新的东西；尚未充分显现的市场或经济潜力；与科学研究的联系日益紧密。总结起来，新兴技术是这样一种技术，它是新出现的、发展速度较快，通常以高科技为基础，可能开辟新的技术和科学领域，具有巨大的市场潜力，可能创造一个新行业或者改变某个老行业，但在现阶段仍然具有不确定性。

表 1 已有研究中对“新兴技术”的定义

时间	机构或研究人员	“新兴技术”的定义
20 世纪 90 年代中期	美国宾夕法尼亚大学[1]	建立在科学基础上的，可能创立一个新行业或改变某个老行业的创新
2005 年	鲁若愚[2]	是一种新观念、新方法、新发明，它以科学为基础，并能够创造一个新的行业或者能够改变一个现有行业并能对经济结构产生重大影响
2005 年	李仕明[3]	具有潜在产业前景和高度不确定性，正在涌现并可能产生巨大变革的技术
2010 年	程跃[7]	那些新近产生甚至正在发展的、对经济结构产生重要影响的高新技术
2012 年	S. Cozzens[5]	快速增长、新兴、具有未开发的市场潜力和高科技基础的技术，这种技术有巨大的潜力，但尚未表现出价值或在业界达成共识
2015 年	A. Breitzman[6]	新兴技术有高速发展的潜力，可能开辟新的技术和科学领域
2015 年	D. Rotolo[8]	一种全新的、相对快速发展的技术，其特点是随着时间的推移，存在一定程度的一致性，并对社会经济领域有相当大的潜在影响，但其最突出的影响在于未来，在出现阶段仍然具有不确定性和模糊性
2018 年	徐建国[4]	知识生产过程中产生的相对快速发展的根本性创新技术，具有影响未来经济和社会发展的潜力

“新兴技术”与“新兴研究领域”“颠覆性技术”“前沿技术”等概念有诸多相似之初，但又存在区别。“新兴研究领域”一般指新兴的科学研究，更多是对新兴科学问题的探索和理论研究，与“新兴技术”相比，其市场应用的要求较低，对经济社会的影响可能较小，但随着科学技术的飞速发展，科学和技术之间的界限越来越模糊，所以一些研究并未对“新兴研究领域”和“新兴技术”进行严格区分，实证分析的领域通常既包含基础研究也包含应用研究和技术研究。与颠覆性技术、前沿技术相比，新兴技术的不确定性更高，其拥有的商业价值只是潜在的，并未充分显现。C. M. Christensen[9]指出，当新的技术创新推翻了市场上现有的主导技术时，被称为“颠覆性技术（Disruptive Technology）”。前沿技术指高技术领域中具有前瞻性、先导性和探索性的重大技术，是未来高技术更新换代和新兴产业发展的重要基础。

2.2 “新兴技术主题识别”及相关概念

技术主题没有明确的定义，不同研究根据各自的研究目的和问题，对其有不同的理解，通常指某技术领域的分支技术领域、技术方向或技术问题，研究的粗细粒度并不一致。在已有的定量研究中，一般用论文或专利的一组关键词/词组或一组文献来揭示技术主题的核心内容[10]。专利文献是技术分析的重要信息源之一，也是研究中最常用的信息源，它集技术信息、法律信息和经济信息于一身，具有新颖、易获取、规范、易检索、时间序列长等特点。随着科学和技术之间的界限越来越模糊，技术主题和研究主题的联系越来越密切，因此也有很多学者同时采用论文和专利文献作为数据源。

与“新兴技术主题识别”相关的概念有“新兴主题监测”（Emerging Topic Detection）、“新兴研究前沿识别”（Emerging Research Fronts Identification）、“新兴趋势探测”（Emerging Trends Detection）、“突发词监测”（Burst Word Detection）和“新事件探测”（New Event Detection）等。这些研究的共同特点是识别或探测最新科学研究活动中已出现但尚未得到广泛认识的新兴话题或主题。“新兴技术主题识别”一般指对技术领域中新出现的分支技术领域、方向或主题的识别，与其它相关研究的识别对象有所不同。

3 新兴技术主题识别方法

科学技术的新兴主题演化、监测和识别一直是政府、企业和科学家感兴趣的研究问题，政府对这方面研究的资助层出不穷。20 世纪 90 年代末，美国国防高级研究计划局（DARPA）实施了“主题监测和跟踪（TDT）计划”，并持续运行了数年[11]。2010 年，《美国竞争

刘小玲，谭宗颖. 新兴技术主题识别方法研究进展[J]. 图书情报工作,2020,64(11):145-152.

法》[12]明确提出将新兴和创新领域的识别作为一项工作目标。2011 年,美国国家情报局局长办公室的情报先进研究计划署(IARPA)资助的"科学展望前瞻计划(FUSE)"[13]旨在开发一套自动化方法,使用科学技术和专利文献中的信息以系统、连续、全面地评估新兴技术。欧盟的 PromTech 项目[14]通过论文文献分析来识别新兴技术。

识别新兴技术或新兴技术主题的方法主要分为两大类,一是以专家主观判断为主的定性研究方法,二是利用文献计量、数据挖掘等对论文、专利文献进行定量研究的方法。定性方法包括德尔菲法、专家头脑风暴法、技术路线图、情景分析、TRIZ 方法等,如欧盟委员会联合研究中心(JRC)的技术预测研究所(IPTS)开发出一种方法(IPTS-TIM)[15],可以通过评价技术的商业化潜力,对现有和未来技术进行识别,支持技术转让过程;F. M. Tseng 等[16]提出将情景规划法、德尔菲法与技术替代模型相结合识别新兴技术;谈毅等[17]结合技术路线图与实物期权方法以识别和选择新兴技术;魏国平[18]利用专家打分法对新兴技术进行了识别。随着信息的爆发式增长和计算机技术的发展,越来越多的学者开始探索基于论文、专利等文献数据的定量分析,为专家的判断提供辅助,弥补专家判断主观性较强的缺陷。本文重点关注新兴技术主题识别的定量研究方法,并根据各方法在分析中所关注的文献特征和属性不同将其分为三类:①主题词或文献统计方法;②引文网络聚类方法;③文本挖掘分析方法。第一类方法主要关注文献关键词/主题词和文献本身的数量特征;第二类方法重在以文献之间的引用关系为基础;第三类则深入文本内容,揭示其语义内涵和关联。三类方法所涉及文献特征和属性的不同使得它们在新兴技术主题识别过程的技术主题抽取、识别指标体系构建和方法有效性验证等环节存在诸多差异。

3.1 主题词或文献统计方法

该类方法通常根据已有的论文、专利分类体系或关键词/词组检索获取技术领域或主题,并以论文、专利文献或其中的主题词/簇对其进行表示,进而根据主题词或文献数量随时间的变化等特征识别出新兴的技术主题。具有代表性的方法是 J. Kleinberg 的突发监测算法[19],J. Kleinberg 使用无限状态自动机对时间序列数据进行建模,时间序列数据状态的转变标志着突发事件的出现,该方法最初用于分析新闻文章等数据流,后来被广泛应用于新兴技术识别的相关研究[20],并已被纳入诸如 Citespace II[21]、SCI2 和 Network Workbench[22]等工具中。M. Bengisu[23]通过关键词/组检索获得材料科学与工程主要分支领域的论文和专利文献,对比各分支领域论文和专利数量随时间的增长,提取出了其中呈快速发展态势的新兴技术领域。E. Schiebel[24]和 I. Roche[14]等结合论文关键词的出现频次、TF-IDF 值、基尼系数将其划分为不寻常的词、既定词和跨领域的词,反映这些词在其他技术领域的扩散情况,以此来识别新兴技术。T. U. Daim 等[25]结合专利分析与情景分析、增长曲线分析方法,对新兴技术进行识别。

3.2 引文网络聚类方法

论文或专利文献之间的引用关系能在一定程度上反映其内容和主题的相关性,该类方法通过对论文或专利的直接引文网络、引文耦合网络或共被引网络进行聚类分析,将内容或主题相近的论文或专利聚集在一起形成一个技术主题,同时利用文献之间的引用关系测度主题的演变路径和趋势,并通过一系列指标识别新兴技术主题。Y. Kajikawa 等[26]对论文文献的直接引文网络进行聚类分析以跟踪能源领域的新兴技术变化,用每个簇中论文的平均出版年作为识别新兴技术主题的指标。H. Small 等[27]结合共被引网络聚类和直接引文网络聚类方法识别出具有新颖性和快速增长的科技主题。J. Hopcroft 等[28]采用引文耦合分析方法识别了计算机领域的若干新兴主题。P. Érdi 等[29]以目标领域中各项专利被其他领域专利引用的情况为基础对目标领域专利集展开聚类分析,提取出其中的子技术集,通过分析这些子技术集在时间维度上的变化,捕捉到新兴技术的出现和发展轨迹。A. Breitzman 等[6]根据"热点专利"间的共被引关系,对"热点专利"及引用"热点专利"的"下一代专利"进行了聚类,并借助专利权人中公共部门比例、科学指数、原创性指数和参考指数等指标对聚类结果进行评价,提取出了其中的新兴技术集合。S. Zhang 等[30]在专利直接引文网络聚类分析基础上,结合网络分析算法进行太阳能光伏领域的新兴技术主题识别。李蓓等[31]依据新兴技术和专利文献的核心特征,建立了基于专利引文耦合聚类的新兴技术识别模型及其相关指标体系,并以美国专利与商标局授权专利数据库为数据源,对纳米技术领域进行了实证分析。

3.3 文本挖掘分析方法

随着数据挖掘和文本分析等计算机技术的发展,越来越多的学者尝试采用这一类方法进行技术发展趋势分析和新兴技术主题的识别,常采用的方法有"主谓

宾”(Subject-Action-Object,简称SAO)结构抽取、向量空间模型、LDA主题模型、机器学习等。J. Kim等[32]通过文本挖掘和决策树的方法进行技术预测,从论文作者、期刊、所属领域,专利的专利权人、所属领域等字段抽取能代表技术主题领域的特征。S. Choi[33]、J. Yoon[34]、李欣[35]和Z. Xiao[36]等基于专利的SAO结构语义分析法识别新兴技术,专利的SAO结构即“主谓宾”结构,可以反映专利技术的功能特征。S. Choi等通过构建名词、动词在SAO结构中的共现网络,基于社会网络分析的节点度数、中心性等指标进行技术发展趋势的分析和新兴技术主题识别。J. Yoon和李欣等通过计算SAO结构的相似度进而获得专利的相似度,以此构建专利网络,再辅以离群节点分析或聚类分析识别新兴技术主题。Z. Xiao等采用主题词簇、SAO结构分析等文本挖掘方法,结合技术路线图和专家判断识别了固体脂质纳米粒子领域的潜在创新和商业应用。任智军[37]、周源[38]、董放[39]等采用LDA模型进行专利技术主题的构建,将一项专利表示为其所属若干主题的概率分布,一个主题表示成若干词的概率分布,在此基础上结合一系列指标和专家判断进行新兴技术主题的识别。K. I. Filippovich等[40]通过机器学习、本体挖掘和实体关联技术进行农业和食品领域的新兴技术识别。P. Yu等[41]利用自组织地图识别新兴技术主题。国内学者王凌燕等[42]构建了工业生物领域的专利高频主题词(题名关键词)共词网络,并进行聚类分析,获得9个技术主题,再通过一系列指标判断新兴技术主题。

4 新兴技术主题识别方法述评

上述三类新兴技术主题识别方法均涉及目标领域确定、数据集构建、技术主题抽取、识别指标体系构建、方法有效性验证等步骤,它们主要在技术主题抽取、识别指标体系构建以及方法有效性验证上存在差异。本文通过构建二维坐标图比较三类方法在主要步骤上的异同,纵坐标表示新兴技术主题识别的三类方法,横坐标表示新兴技术主题识别的三个主要步骤(见图1)。三类方法在技术主题抽取上的区别最为明显,在识别指标体系构建上各有侧重,但也有共同采用的指标,在方法有效性验证方面共性最大,三类方法在各识别步骤上各有优劣,笔者将分别进行分析和展望。

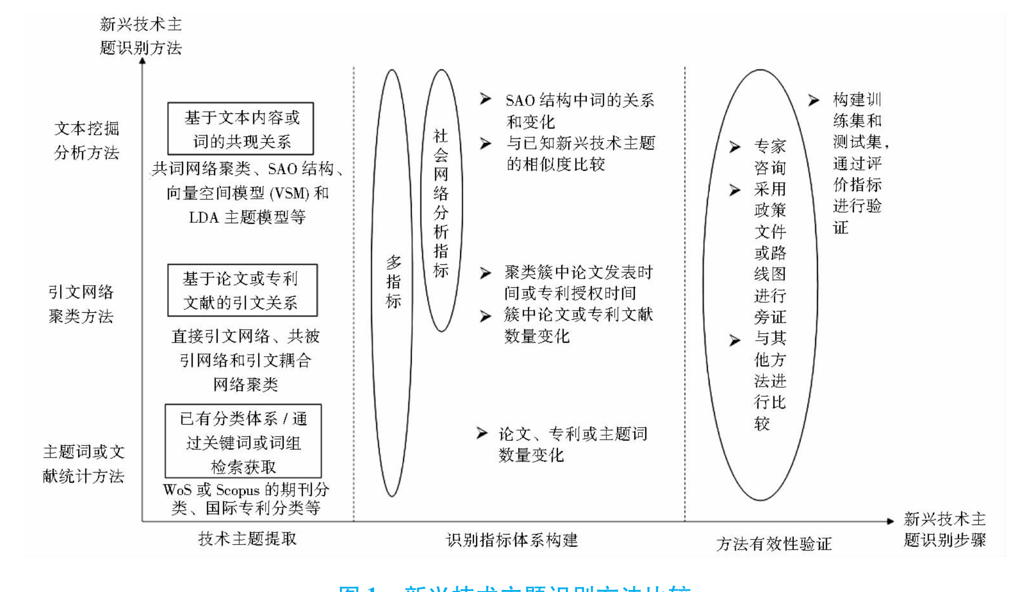


图1 新兴技术主题识别方法比较

4.1 技术主题抽取

主题词或文献统计方法主要基于已有分类体系或关键词/词组检索获得论文或专利数据的主题划分,已有分类体系包括Web of Science(WoS)或Scopus数据库的论文期刊分类、国际专利分类等,该方法采用通用的分类方法,易获得认可,但难以反映科学研究和技术发展的动态变化。引文网络聚类方法主要根据论文或专利之间的引用关系构建技术主题,包括基于直接引用关系、共被引关系和引文耦合关系三种[43-44],该方法与基于已有分类体系的方法相比,揭示了单篇文献之间的关系,能够反映技术的动态变化,但也存在一些缺陷,如引用的动机具有多样性,有引用关系的文献之间并不一定具有主题上的相似性,而且引用发生在文献发表之后,存在时滞问题。文本挖掘分析方法通常根据文本内容或词之间的共现关系构建技术主题,如上文中提到的共词网络聚类方法、SAO结构、向量空间

模型(VSM)、LDA 主题模型等,这一类方法基于目前快速发展的数据挖掘、深度学习等技术,对文本内容进行深度揭示,能够更准确地抽取技术主题,但也有不断完善的空间。向量空间模型基于词频进行计算,但词频难以准确反映词的语义和词间关系,以此构建的向量也难以准确地测度文本的主题内容。目前有一些研究对向量空间模型进行改进,主要是采用外部词典,如 WordNet 等,对词的语义相似度进行度量,并结合 TF-IDF 算法进行文本表示和分类,这种方法仍然难以根据词的上下文信息准确度量词的含义[45－46]。LDA 主题模型[47]方法是用来在一系列文档中发现抽象主题的一种统计模型,把每篇文档表示成所属主题的概率分布,而每个主题表示成一组词的概率分布,但概率分布表示仅仅描述语料中的共现统计关系,并不是文本特征表示的最好选择,通常难以从一组词判断出确切的主题含义。

4.2 新兴技术主题识别指标体系构建

在识别指标体系构建上,三类新兴技术主题识别方法各有侧重。主题词或文献统计方法通常采用论文或专利文献、主题词的数量变化指标,如 M. Bengisu 等[23]采用论文和专利数量变化指标,P. Érdi[29]、E. Schiebel 等[24]基于关键词的出现频次、TF-IDF 值和基尼系数,这些指标本质上反映了文献和主题词数量的变化。引文网络聚类方法采用的指标更加多样,如聚类簇中论文发表时间或专利授权时间、簇中论文或专利数量的变化等,Y. Kajikawa 等[26]利用主题簇中论文的平均出版年作为识别能源领域新兴技术的指标,H. Small 也采用了类似指标。文本挖掘分析方法则采用更多能够揭示文本内容的指标,如 SAO 结构中词的关系和变化、与已知新兴技术主题的相似度比较等。由于引文网络聚类方法和文本挖掘分析方法能够根据文献的引用关系或内容相关关系构建技术主题的关系网络,因此也采用社会网络分析的一些指标,如点度中心度、中介中心度、结构洞等,S. Zhang[30]和王凌燕等[42]采用了这一类指标。有少数学者基于多指标进行新兴技术主题识别,如 J. Kim 等[32]采用论文作者数量、期刊、专利权人数量、论文或专利所属领域等指标;A. Breitzman 等[6]采用专利权人类型、技术与科学的关联、技术原创性指数、引用前人技术情况等指标;李蓓等[31]采用簇中专利授权时间中位数和专利权利要求数量指标。通过以上分析可知,现有研究中,大多指标体系对新兴技术主题特征的反映不够全面,有进一步改进和完善的空间。

4.3 方法有效性验证

三类方法均采用专家咨询、利用政策文件或路线图等进行旁证或与其它方法进行比较等对识别方法和指标体系的有效性进行验证,文本挖掘分析方法则开始探索通过构建训练集和测试集,采用相关评价指标进行更加严格的验证,但目前该类研究数量仍较少。专家咨询方法的缺点是专家带有一定的主观性,也受专家的知识范围的影响。有学者利用政策文件或路线图等进行旁证,如 Y. Kajikawa 等[26]将识别出的能源领域新兴技术与日本政府机构绘制的该领域专家路线图进行对比,这种方法也存在所识别新兴技术主题的粗细粒度不同,从而导致与路线图不能完全匹配的问题,且在评价时依赖人工解读。与已有研究方法进行对比也是采用较多的验证方法,如 Q. Wang[48]将识别出的新兴主题与已有文献中提及的新兴主题进行对比,该方法可能存在的问题是已有研究和本研究对新兴主题的定义有所不同,识别的主题粗细粒度也可能不同。

4.4 新兴技术主题识别方法展望

基于论文或专利文献的技术主题抽取的准确性依赖于对文献内容的准确理解和分析,而目前快速发展的数据挖掘、深度学习等技术可以应用于该问题的研究。近年来,一些基于深度学习的自然语言处理模型在文本语义分析上取得了较好效果,如神经网络语言模型以及 Google 公司在 2013 年推出的 Word2Vec 模型通过学习分布式词向量对文本进行表示,利用了词的上下文信息,可以解决数据稀疏、缺失语义表达能力等问题,能够在一定程度上解决共词网络聚类、向量空间模型、LDA 主题模型等方法不能准确反映词含义的问题,因此可以探索这类方法在技术主题抽取上的应用。

新兴技术具有新出现、发展速度快、以高科技为基础、市场潜力巨大等特征,现有研究的新兴技术主题识别指标通常考虑不够全面,仅根据其中一个或几个特征构建指标,笔者认为应更加全面、系统地考虑可用于新兴技术主题识别的指标,再通过一些方法对指标进行遴选,从而优化新兴技术主题识别效果。基于新兴技术主题的内涵和特征,并结合现有研究,笔者构建了包含以下 7 个特征维度的指标体系:新颖性、规模、增长速度、影响力、科学关联、市场潜力、不确定性(见表 2),在以后的研究中将通过实验进行指标遴选和评价。

方法的有效性验证方面,有监督的机器学习常用来研究分类问题,新兴技术主题的识别本质上也是一种分类问题,即把技术主题集合划分为新兴技术主题

和非新兴技术主题。因此可以考虑事先构建新兴技术主题识别的训练集,以此对识别指标进行遴选,再通过测试集验证指标和方法的有效性。尽管目前已有少量研究采用这类方法进行验证,但还处于探索阶段,今后仍有较大发展空间。

表 2 新兴技术主题识别指标体系及其含义

特征	指标	含义
新颖性[48-49]	新词/词组数的增长率	新词/词组的出现和增长代表新概念、方法、工具的出现
	专利的平均授权年	平均授权年越大,说明专利簇所代表的技术越新
	施引专利的平均授权年	施引专利的授权年越大,表明该技术近年来越受关注
	参考专利/论文的平均授权年/出版年	参考专利或论文的授权年或出版年越大,表明该技术引用的技术或科学原理越新,更具新颖性
规模[13,50]	专利数量	专利数达到一定规模的技术更有可能是新兴技术
	专利权人数量	参与研发的专利权人达到一定数量的技术更有可能是新兴技术
增长速度[5,51]	专利数量每一年的增长率	专利数量增长快,表明该技术处在快速发展阶段
	专利权人每一年的增长率	参与到技术研发中的企业或科研机构越来越多,也表明该技术正快速发展
影响力[52,30]	专利篇均被引频次	被引频次高,表明该技术的影响力大
	专利的平均权利要求数	权利要求数大,表明专利保护的技术点多,质量高
科学关联[51,18]	专利平均参考论文数	参考的论文多,表明该技术与科学的关联性强,反映新兴技术的高科技特征
市场潜力[53-54]	专利的转化比例	新兴技术具有较大市场潜力,体现在与其他技术相比,专利转化率较高
不确定性[8]	专利权人中公共机构所占比例	新兴技术具有不确定性,因此政府投入较多,企业投入相对较少

此外,新兴技术主题识别的定量研究方法往往基于某领域的小批量数据集,而随着学科之间的交叉融合越来越多,新兴技术很可能出现在多个学科的交叉领域。因此,从多学科领域的大批量数据集中发现新兴技术具有重要意义。调研发现基于大数据集进行分析的研究较少,H. Small 等[27] 和 ERACEP 项目[55] 基于一段时间内全领域的论文数据识别出新兴研究主题,并非新兴技术主题的识别。目前,大多数新兴技术主题的识别是对预先确定领域的回溯性分析,而非侧重在识别新兴技术的方法学研究,一般把具有突发可能性的主题作为实证研究数据,再通过一定方法验证该主题具有突发性,或从中发现突发特征最明显的子主题,严格来说,这一类研究并非真正意义上的新兴技术主题的识别。

5 结语

通过对国内外新兴技术主题识别研究的系统调研和综合分析,本文将新兴技术主题识别的定量研究方法分为三类:主题词或文献统计方法;引文网络聚类方法;文本挖掘分析方法。这些方法在技术主题抽取、识别指标体系构建、方法有效性验证等方面存在差异,它们各有优点,但也都存在一定缺陷和不足。随着深度学习等技术的发展,在论文和专利等文本内容的准确解析、技术主题抽取等方面存在的问题可以得到更好的解决,未来可以探索一些基于深度学习的自然语言处理模型在技术主题抽取上的应用。此外,还需要在对"新兴技术主题"的内涵进行深入理解的基础上,构建较为完善的识别指标体系,并构建新兴技术主题的训练集和测试集,借助有监督的机器学习方法对新兴技术主题训练集进行学习,遴选真正相关的指标,再通过测试集对指标体系和方法的有效性进行更严格的验证。

参考文献:

[1] 乔治·戴,保罗·休梅克,石莹. 沃顿论新兴技术管理[M]. 北京:华夏出版社, 2002.
[2] 鲁若愚, 张红琪. 基于快变市场的新兴技术产品更新策略[J]. 管理学报, 2005(3):67-70.
[3] 李仕明, 肖 磊, 萧延高. 新兴技术管理研究综述[J]. 管理科学学报, 2007, 10(6): 12.
[4] 徐建国, 李孟军, 游翰霖. 新兴技术识别研究进展[J]. 情报杂志, 2018, 37(12):11-16.
[5] COZZENS S, GATCHAIR S, KANG J, et al. Emerging technologies: quantitative identification and measurement[J]. Technology analysis & strategic management, 2010, 22(3): 361-376.
[6] BREITZMAN A, THOMAS P. The Emerging Clusters Model: a tool for identifying emerging technologies across multiple patent systems[J]. Research policy, 2015, 44(1): 195-205.
[7] 程跃, 银 路. 基于企业动态能力的新兴技术演化模型及案例研究[J]. 管理学报, 2010, 7(1): 43-49.
[8] ROTOLO D, HICKS D, MARTIN B R. What is an emerging technology? [J]. Research policy, 2015, 44(10): 1827-1843.
[9] CHRISTENSEN C M. The innovator's dilemma[M]. Boston: Harvard Business Review Press,1997.
[10] 胡阿沛, 张静, 雷孝平, 等. 基于文本挖掘的专利技术主题分析研究综述[J]. 情报杂志, 2013, 32(12): 88-92.

[11] ALLAN J, CARBONELL J G, DODDINGTON G, et al. Topic detection and tracking pilot study final report[R]. Virginia: Defense Advanced Research Projects Agency, 2000.
[12] United States Congress. America COMPETES Act [EB/OL]. [2019－09－13]. https://www. congress. gov/111/plaws/publ358/PLAW-111publ358. pdf.
[13] Intelligence Advanced Research Projects Activity. Foresight and Understanding from Scientific Exposition (FUSE) [EB/OL]. [2019－09－13]. https://www. iarpa. gov/index. php/research-programs/fuse.
[14] ROCHE I, BESAGNI D, FRANÇOIS C, et al. Identification and characterisation of technological topics in the field of Molecular Biology[J]. Scientometrics, 2010, 82(3): 663－676.
[15] MONCADA-PATERNÒ-CASTELLO P, ROJO J, BELLIDO F, et al. Early identification and marketing of innovative technologies: a case study of RTD result valorisation at the European Commission's Joint Research Centre[J]. Technovation, 2003, 23(8): 655－667.
[16] TSENG F M, CHENG A C, PENG Y N. Assessing market penetration combining scenario analysis, Delphi, and the technological substitution model: the case of the OLED TV market[J]. Technological forecasting and social change, 2009, 76(7): 897－909.
[17] 谈毅，黄燕丽. 基于过程的新兴技术规划与选择模型研究[J]. 科技管理研究，2007，27(8)：5－8.
[18] 魏国平. 新兴技术管理策略研究[D]. 杭州：浙江大学，2006.
[19] KLEINBERG J. Bursty and hierarchical structure in streams[J]. Data mining and knowledge discovery, 2003, 7(4): 373－397.
[20] MANE K K, BÖRNER K. Mapping topics and topic bursts in PNAS[J]. Proceedings of the National Academy of Sciences, 2004, 101(S1): 5287－5290.
[21] CHEN C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the Association for Information Science and Technology, 2006, 57(3): 359－377.
[22] BÖRNER K, HUANG W, LINNEMEIER M, et al. Rete-netzwerk-red: analyzing and visualizing scholarly networks using the Network Workbench Tool[J]. Scientometrics, 2010, 83(3): 863－876.
[23] BENGISU M. Critical and emerging technologies in materials, manufacturing, and industrial engineering: a study for priority setting[J]. Scientometrics, 2003, 58(3): 473－487.
[24] SCHIEBEL E, HÖRLESBERGER M, ROCHE I, et al. An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices[J]. Scientometrics, 2010, 83(3): 765－781.
[25] DAIM T U, RUEDA G, MARTIN H, et al. Forecasting emerging technologies: Use of bibliometrics and patent analysis[J]. Technological forecasting and social change, 2006, 73(8): 981－1012.
[26] KAJIKAWA Y, YOSHIKAWA J, TAKEDA Y, et al. Tracking emerging technologies in energy research: toward a roadmap for sustainable energy[J]. Technological forecasting and social change, 2008, 75(6): 771－782.
[27] SMALL H, BOYACK K W, KLAVANS R. Identifying emerging topics in science and technology[J]. Research policy, 2014, 43(8): 1450－1467.
[28] HOPCROFT J, KHAN O, KULIS B, et al. Tracking evolving communities in large linked networks[J]. Proceedings of the National Academy of Sciences, 2004, 101(S1): 5249－5253.
[29] ÉRDI P, MAKOVI K, SOMOGYVÁRI Z, et al. Prediction of emerging technologies based on analysis of the US patent citation network[J]. Scientometrics, 2013, 95(1): 225－242.
[30] ZHANG S, HAN F. Identifying emerging topics in a technological domain[J]. Journal of intelligent & fuzzy systems, 2016, 31(4): 2147－2157.
[31] 李蓓，陈向东. 基于专利引用耦合聚类的纳米领域新兴技术识别[J]. 情报杂志，2015 (5)：35－40.
[32] KIM J, HWANG M, JEONG D H, et al. Technology trends analysis and forecasting application based on decision tree and statistical feature analysis[J]. Expert systems with applications, 2012, 39(16): 12618－12625.
[33] CHOI S, YOON J, KIM K, et al. SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells [J]. Scientometrics, 2011, 88(3): 863－883.
[34] YOON J, KIM K. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection[J]. Scientometrics, 2012, 90(2): 445－461.
[35] 李欣，王静静，杨梓，等. 基于 SAO 结构语义分析的新兴技术识别研究[J]. 情报杂志，2016，35(3)：80－84.
[36] XIAO Z, LU H, ALAN P, et al. Tracing the system transformations and innovation pathways of an emerging technology: solid lipid nanoparticles[J]. Technological forecasting and social change, 2018, 146(2): 785－794.
[37] 任智军，乔晓东，张江涛. 新兴技术发现模型研究[J]. 现代图书情报技术，2016，32(7):60－69.
[38] 周源，刘宇飞，薛澜. 一种基于机器学习的新兴技术识别方法:以机器人技术为例[J]. 情报学报，2018，37(9):83－99.
[39] 董放，刘宇飞，周源. 基于 LDA-SVM 论文摘要多分类新兴技术预测[J]. 情报杂志，2017(7):44－49,137.
[40] FILIPPOVICH K I, DENISOVICH B P, STANISLAVOVNA L A. Big-data-augmented approach to emerging technologies identification: case of agriculture and food sector [J]. Social science research network, 2017(1): 130－134.
[41] YU P, LEE J H. A hybrid approach using two-level SOM and combined AHP rating and AHP/DEA-AR method for selecting optimal promising emerging technology[J]. Expert systems with applications, 2013, 40(1): 300－314.
[42] 王凌燕，方曙，季培培. 利用专利文献识别新兴技术主题的技术框架研究[J]. 图书情报工作，2011，55(18)：74－23.

[43] FURUKAWA T, MORI K, ARINO K, et al. Identifying the evolutionary process of emerging technologies: a chronological network analysis of World Wide Web conference sessions[J]. Technological forecasting and social change, 2015, 91: 280 – 294.
[44] ZHANG Y, PORTER A L, HU Z, et al. "Term clumping" for technical intelligence: a case study on dye-sensitized solar cells [J]. Technological forecasting and social change, 2014, 85(2): 26 – 39.
[45] PATIL L H, ATIQUE M. A novel approach for feature selection method TF-IDF in document clustering[C]//Advance Computing Conference (IACC), 2013 IEEE 3rd International. Ghaziabad: IEEE, 2013: 858 – 862.
[46] 黄承慧, 印鉴, 侯昉. 一种结合词项语义信息和 TF-IDF 方法的文本相似度量方法[J]. 计算机学报, 2011, 34(5): 856 – 864.
[47] YAN E. Research dynamics: measuring the continuity and popularity of research topics[J]. Journal of Informetrics, 2014, 8(1):98 – 110.
[48] WANG Q. A bibliometric model for identifying emerging research topics[J]. Journal of the Association for Information Science and Technology, 2017, 69(2): 290 – 304.
[49] GHO H, WEINGART S, BORNER K. Mixed-indicators model for identifying emerging research areas[J]. Scientometrics, 2011, 89 (1):421 – 435.
[50] CHRISTOPHER E, SILBERGLITT R. Identification and analysis of technology emergence using patent classification[EB/OL]. [2019 – 09 – 01]. https://www.rand.org/pubs/research_reports/RR629.html.
[51] 郭瑞兰. 情景规划在新兴技术战略制定中的应用[D]. 成都: 电子科技大学, 2007.
[52] NOH H, SONG Y, LEE S. Identifying emerging core technologies for the future: case study of patents published by leading telecommunication organizations[J]. Telecommunications policy, 2016, 40(10/11):956 – 970.
[53] 张伟, 陈绍刚. 新兴技术采用的生命周期各阶段技术风险研究[J]. 科技管理研究, 2007, 27(5):161 – 163.
[54] 黄鲁成. 基于属性综合评价系统的新兴技术识别研究[J]. 科研管理, 2009,30(4):190 – 194.
[55] European Research Council. ERACEP-emerging research areas and their coverage by ERC-supported projects final report[EB/OL]. [2019 – 09 – 13]. https://erc.europa.eu/sites/default/files/document/file/Eracep_final_report.pdf.

作者贡献说明:
刘小玲:参与论文框架设计、撰写论文初稿、修改论文、定稿;
谭宗颖:提出研究思路与论文框架、论文审核与修改。

Progress on Methods of Emerging Technology Topics Identification

Liu Xiaoling[1,2] Tan Zongying[1]
[1] National Science Library, Chinese Academy of Sciences, Beijing 100190
[2] Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100049

Abstract: [**Purpose/significance**] Identification of emerging technology topics not only can contribute to track the development of technologies, but also can capture the future development opportunities and trends of technologies. Reviewing the quantitative methods of emerging technology topics identification and making a comparison of them can provide reference for an improvement of the methods. [**Method/process**] Firstly, concepts such as "emerging technology" and "emerging technology topics identification" were analyzed; then qualitative and quantitative research methods of emerging technology topics identification at home and abroad were investigated, focusing on bibliometrics and data mining. Quantitative methods were divided into 3 categories: keywords or documents statistical method, citation network clustering and text mining. Similarities, differences and shortcomings of above methods in the extraction of technology topics, construction of emerging technology topic identification indictors, methods verification were analyzed. Improvement methods are provided preliminarily. [**Result/conclusion**] The three types of methods have their own characteristics, advantages and disadvantages in the three steps of emerging technology topics identification, and there is room for further improvement. In the future, we can explore the use of techniques such as deep learning to identify technology topics accurately, and build a group of more comprehensive and systematic emerging technology topic identification indicators, as well as more rigorous method validation based on machine learning.

Keywords: emerging technology topics identification methodology research bibliometrics text mining